

J. Perinat. Med.
16 (1988) 345

From mainframe to micro: decentralization of perinatal epidemiology

Sidney F. Bottoms, Glenn Tse, and Larry Chik

Wayne State University, Hutzel Hospital, Detroit, Michigan, U.S.A.

1 Methods

The 1980 National Natality Survey is a weighted, stratified, randomized, national sample of all births occurring in the United States during 1980 [2]. It includes 2641 bytes of mainly coded information for each of 9941 live births. The total dataset is about 25 megabytes, and is usually supplied on tape along with the 1980 National Fetal Mortality Survey. Documentation described the definition, tape location (offset in bytes from the beginning of each record), length and source for each variable [1]. The tape also provided support for variable selection and analysis using the Table Processing Language, a program available for mainframe computers. Cases of infant mortality were entered into the computer separately and then linked to the National Survey data.

Tapes were read into the Department of Obstetrics and Gynecology minicomputer, a Digital Equipment Corporation Vax® 11/750 operating on Berkeley 4.1 Unix®. The 2641 byte record length far exceeded the 512 byte maximum line length configuration of the minicomputer. Although the minicomputer was capable of handling records this long, a short C language program was written to divide each record into shorter segments so that line-oriented operating system utilities could be used more conveniently. The Unix utility "compact" was used to compress files and minimize disk storage space requirements. A short Unix shell program using the "awk" text scanning and processing utility was developed to support variable selection and extraction of working datasets. The shell program also automatically generated the instruction set for preliminary descriptive statistical analysis. Documentation was facilitated by linking the tape location supplied with the original

Curriculum vitae

Dr. BOTTOMS completed medical school and residency at the University of Florida. He subsequently completed fellowship training in Maternal-Fetal Medicine at Cleveland Metropolitan General Hospital and remained on the faculty of Case Western Reserve University. He is currently Associate Professor of Obstetrics and Gynecology and Director of Maternal-Fetal Medicine at Wayne State University in Detroit.



documentation to the variable name used in subsequent processing. BMDP® programs were used for statistical analysis.

An ASCII-based compression scheme employing a stream editor was used to abbreviate transfer of the dataset to a Wyse PC286® microcomputer, an IBM AT® compatible device. Microport System V Unix® was used as the operating system for processing the complete dataset; MSDOS® was the operating system employed during statistical analysis of working datasets. The complete dataset was transferred to the microcomputer by phone using Kermit, telecommunications software developed at Columbia University, to facilitate data integrity. A shell program similar to the one described above was developed, taking advantage of newer utilities available in System V Unix ("pack" for file compression, "cut" and "paste" for data extraction). SYSTAT® programs were used for statistical analysis.

2 Results

Minicomputer and microcomputer implementations were validated and compared on a practical basis by performing and identical analytic sequence on each machine. A total of 55 variables (of which about half were missing value indicators) were retrieved from the 9941 live births, data sufficient for several related studies. Cases were then limited to the 7825 births in wedlock, one new variable was derived, and the new dataset was saved. Finally, eight contingency table analyses were performed. Counts, ranges and descriptive statistics from both smaller computers matched published values derived on mainframes [1–3]. Counts and ranges were identical were identical on the minicomputer and the microcomputer. Means, standard deviations and standard errors agreed within at least 6 digits or 3 decimal places.

A comparison of the times required for the analytic sequence is presented in table I. The minicomputer runs were made during a period of light usage (the load average ranged from 1.5–2.0); no other jobs were being run on the microcomputer. These times are also not benchmarks for the computers themselves because different operating systems, shell programs and statistical packages were used. It should be noted, however, that operating systems and applications software available for minicomputers and microcomputers are usually different, a factor to be considered in any practical comparison. Overall, the microcomputer implementation compared favorably to that of the minicomputer. However, this advantage could be lost if more extensive statistical analysis were required.

The results of the analytic sequence itself are depicted in figure 1. Infant mortality rates were tabulated according to the amount alcohol consumed at a time and different cutpoints were used for a series of contingency table analyses. The results suggest a threshold for alcohol consumption associated with infant mortality.

Table I. Time in minutes required to accomplish identical analytical sequences

	Mini-computer	Micro-computer
Retrieve 55 variables	178	48
Select, describe and save cases	16	52
Tabular analysis	4	10
TOTAL	198	110

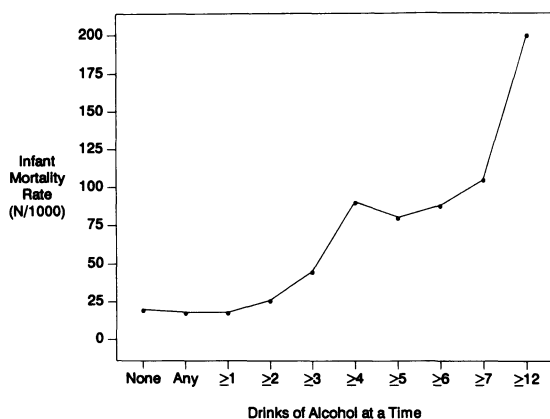


Figure 1. The association of infant mortality with maternal alcohol consumption among the 7825 married women in the 1980 United States National Natality Survey.

3 Discussion

The possibility of analyzing large perinatal datasets on small computers has been demonstrated in this project. Whether this approach is practical is a more complicated issue that will be discussed by considering the major limiting factors. In contrast to mainframes, programs to manage such large datasets conveniently are not readily available for smaller machines. In theory it would be possible to write a conventional program to retrieve each working dataset, but shortage of skilled personnel and/or cost would be prohibitive. Minimizing requirements for personnel time and expertise by generous use of powerful operating system utilities seemed to be a key to the success of this project. Although DOS utilities are of little help, the availability of Unix makes managing large datasets on microcomputers more practical.

Two other limiting factors bear mentioning. Requirements for disk storage and backup were greatly reduced by file compression programs (available on DOS as well as Unix) and 1.2 megabyte floppy disks. Since the entire dataset could be stored on only seven floppy disks, a tape drive or larger hard disk for the microcomputer was not necessary. File transfer has become less of a limitation with the advent of telecommunications programs with extensive error checking. The final considerations are of application, rather than development. Installing and maintaining a more complicated operating system will be impractical

for many personal computer users. Once installed, it is apparent that run times on either microcomputer or minicomputer preclude interactive use; batch programming and more careful planning are necessary. Expertise in study design and statistical analysis are also needed. These limitations must be weighed against the alternatives, however. Careful study design and statistical analysis are required regardless of how the data are collected. If the research question can be addressed using a

source such as the National Natality Survey, performing a chart review would probably take for longer and be less likely to provide a reliable answer.

We conclude that microcomputer analysis of large perinatal datasets is not only possible, but may also be practical. This approach also appears to be an advantageous alternative to some chart review studies.

Abstract

To evaluate the feasibility of analyzing large perinatal datasets on smaller computers, the 1980 National Natality Survey was downloaded first to a minicomputer and then to a microcomputer. Operating system utilities were used to minimize programming and storage requirements. Accuracy was confirmed by comparing descriptive statistics with published values derived on

mainframes. Based on a typical analytic problem, implementation on the microcomputer compared favorably to implementation on the minicomputer. Our results suggest that the microcomputer may be a practical alternative to the mainframe for perinatal epidemiological analysis.

Keywords: Computer operating system, epidemiology, fetal alcohol effects, infant mortality, microcomputer, minicomputer, perinatology.

Zusammenfassung

Vom Groß- zum Mikrorechner: Dezentralisation der perinatalogischen Epidemiologie

Das Ziel dieser Studie war die Untersuchung der Möglichkeit, große perinatalogische Datenmengen auf Klein-/ Mikrocomputern auszuwerten. Der Nationale Geburtenbericht von 1980, der für jede der 9941 Geburten 2641 Bytes an Informationen enthält, wurde als Ausgangsmaterial in einen Vax 11/750 Kleincomputer geladen. Anhand von Benutzerrouninen des Unix Betriebssystems wurden Variablenselektion, Dokumentation und Datenverdichtung vorgenommen. Der gesamte Datensatz wurde dann per Fernübertragung auf einen IBM AT-kompatiblen Mikrorechner mit Unix und DOS Betriebssystem überspielt.

Die Rechnungen auf dem Klein- bzw. Mikrorechner wurden dadurch einander gegenübergestellt, daß von 7825 ehelich geborenen Kindern je 55 Variablen übernommen, eine neue Variable berechnet, der neue Daten-

satz abgelegt und 8 Kontingenztafeln berechnet wurden. Die Anzahlen, die Variationsbreiten und beschreibende Statistiken entsprechen den veröffentlichten Werten von Großrechnern. Eine Gegenüberstellung für Sequenzanalysen (Tabelle I) zeigte, daß den Mikrorechnern vergleichsweise zu den Kleinrechnern ein Vorteil zukommt. Die Ergebnisse der Sequenzanalysen zeigen, wie in Abb. 1 dargestellt, daß in bezug auf die Kindersterblichkeit dem Alkoholkonsum ein Schwellenwert zugeordnet werden kann.

Leistungsfähige Betriebssystemroutinen, Datenverdichtung und 1.2 Megabyte Disketten erwiesen sich als Schlüssel zum Erfolg. Wir schließen daraus, daß Mikrorechneranalysen großer perinatalogischer Datensätze nicht nur möglich, sondern auch zweckmäßig sind. Darüber hinaus erscheint dieses Vorgehen als eine echte Alternative zu einigen graphischen Übersichtsstudien.

Schlüsselwörter: Computer Betriebssystem, Epidemiologie, fetale Alkoholauswirkungen, Kindersterblichkeit, Kleinrechner, Mikrorechner, Perinatalogie.

Résumé

De l'ordinateur central à la microinformatique: décentralisation de l'épidémiologie périnatale

Le but de cette étude a été d'évaluer la faisabilité de l'analyse de l'ensemble des données périnatales dans un environnement mini/micro informatisé. L'enquête nationale de natalité de 1980, qui consiste en 2641 bytes d'information pour chacune des 9941 naissances, a été

primitivement entrée dans un miniordinateur VAX 11/750. On a utilisé les services de système Operating UNIX pour aider la sélection variable, le documentation et la compression de déroulement. Ensuite la totalité de l'ensemble des données a été télécommuniquée à un microordinateur IBM AT compatible équipé de systèmes opérateurs UNIX et DOS.

On a comparé les réalisations du miniordinateur et du microordinateur en reprenant 55 variables, en limitant le nombre de naissances aux 7825 couplées, en dérivant une nouvelle variable, en mettant de côté le nouvel ensemble de données et en réalisant huit analyses de tableau de contingence. La comparaison des temps nécessaires pour les séquences analytiques (tableau I) ont montré que la réalisation du microordinateur est favorable par rapport au miniordinateur. Les résultats de la séquence analytique sont décrits dans la figure 1,

ils suggèrent un seuil pour la consommation d'alcool accompagnée d'une mortalité infantile.

Les services de Système Opérateur puissant, la compression de déroulement et les disquettes de 1,2 mégabytes semblent être des clés du succès. Nous concluons que l'analyse par microordinateur d'un ensemble de données périnatales vastes est non seulement possible mais qu'elle est également pratique. Cette approche également apparaît être une alternative avantageuse aux quelques études de revue de diagramme.

Mots-clés: Effets de l'alcoolisme fœtale, épidémiologie, microordinateur, miniordinateur, mortalité infantile, ordinateur, périnatalogie, système Opérateur.

References

- [1] National Center for Health Statistics: Public Use Tape Documentation, National Natality Survey, 1980, National Fetal Mortality Survey, 1980. United States Department of Health and Human Services, Public Health Service, National Center for Health Statistics. Hyattsville, Maryland 1984
 - [2] PLACEK PJ: The 1980 National Natality Survey and National Fetal Mortality Survey — methods used and PHS agency participation. Pub Health Rep 99 (1984) 111
 - [3] PRAGER K, H MALIN, D SPIEGLER, PV NATTA, PJ PLACEK: Smoking and drinking behavior before and during pregnancy of married mothers of live-born infants and stillborn infants. Pub Health Rep 99 (1984) 117
- Sidney F. Bottoms, M. D.
Department of Obstetrics & Gynecology
Hutzel Hospital
4707 St. Antoine Boulevard
Detroit, MI 48201
U. S. A.